1.

Ripple joins for online aggregation

Online aggregation

Nested loop join, relational algebra

Ripple join：a. Block

b.Hash

math tool: central limit theorem for confidence interval

Trade off between rate of update and shrinkage of the confidence interval:

Set up the rate, then optimize the shrinkage of interval

2. A cost based query optimizer for online processing of SQL queries

Cost function: the object of the function

1. Specific relation on some node: division of the relation

The cost based query optimizer, the optimization variables

Input: query, related operators, available nodes, cost function

Output:

Query plan: topology relation, operator distribution on node,

3. A scalable hash ripple join algorithm

1. Memory overflow: how does the memory overflow result in the transformation from hash ripple to block ripple?
2. Support online aggregation:

Centralized two phase aggregation,

Only one layer divide and conquer

4. relational join

(1) reduce side join

(2) map side join

(3)memory backed join